Journal of Network Intelligence ©2025 ISSN 2414-8105 (Online)
Taiwan Ubiquitous Information Volume **10**, Number **4**, November **2025**

Research on Trajectory Detection and Repair Based on ETC Data

Song-Yang Wu*

School of Computer Science and Mathematics
Fujian University of Technology, Fuzhou, 350118, China
2361433667@qq.com

Fu-Min Zou

School of Electronic, Electrical Engineering and Physics
Fujian University of Technology,Fuzhou, 350118, China
fmzou@fjut.edu.cn

Zhao-Yi Zhou

School of Electronic, Electrical Engineering and Physics
Fujian University of Technology, Fuzhou, 350118, China
1911181484@qq.com

Ting Ye

Fujian Key Laboratory of Automotive Electronics and Electric Drive
Fujian University of Technology, Fuzhou, 350118, China
1402702966@qq.com

*Corresponding author: Song-Yang Wu

Received April 30, 2023, revised April 15, 2024, accepted March 17, 2025.

Abstract. *The ETC system generates a massive amount of transaction data every day, recording the traffic information of the vast majority of vehicles. Moreover, ETC transaction data has the characteristics of high dimensionality and rich content, making it the cornerstone of intelligent transportation data. But the difference between ETC systems and traditional IoT systems lies in their diverse types of devices, harsh working environments, and short detection time, which leads to issues with transaction data. Therefore, there are also corresponding issues with the construction of vehicle trajectory data based on ETC transaction data, and the research on anomaly detection and repair of these trajectory data is not sufficient, making it difficult to effectively carry out reasonable and complete repair work. How to scientifically and effectively detect and repair the anomaly of track data has become the key problem that ETC data governance needs to solve. This article conducts anomaly detection and repair work on vehicle trajectories based on ETC transaction data, and proposes a repair method for trajectory data. The proposed method is used to repair a day's trajectory data.*

Keywords: Expressway, ETC big data, Trajectory detection, Abnormal data repair

1. **Introduction.** The ETC system generates a massive amount of transaction data every day, recording the traffic information of the vast majority of vehicles. Moreover, ETC transaction data has the characteristics of high dimensionality and rich content, making it the cornerstone of intelligent transportation data. But the difference between ETC systems and traditional IoT systems lies in their diverse types of devices, harsh working environments, and short detection time, which leads to issues with transaction data. The

vehicle trajectory data constructed based on ETC transaction data also has corresponding problems, and the research on anomaly detection and repair of these trajectory data is not sufficient, making it difficult to effectively carry out reasonable and complete repair work.

In terms of anomaly detection, Yang et al. [1] proposed an anomaly trajectory detection and calculation method based on the TRASMIL framework by dividing each motion trajectory into independent sub trajectories and proposing a granularity metric with diversity to measure the quality of the divided sub trajectories. Then, a sequence learning model was used to model the sub trajectories; Braei et al. [2] studied 20 methods for anomaly detection in univariate time series data. By analyzing the accuracy and efficiency of each method, they provided different detection methods suitable for different types of anomaly data; Pang et al. [3] conducted a comprehensive analysis of the objective function, advantages and disadvantages involved in anomaly detection based on deep learning, and proposed relevant solutions to solve future new anomaly detection problems; Hu et al. [4] established an abnormal driving detection model based on deep learning for abnormal driving behavior, used stacked sparse self encoder model to learn general driving behavior characteristics, and added a denoising method to increase the robustness of feature expression, combined with dropout technology to avoid overfitting in the whole process. The proposed model has better performance in abnormal driving detection; Zhao et al. [5] proposed an abnormal trajectory detection method called TADSS, which measures the time, velocity, and position feature values of trajectory data through three kernel functions. The semantic features of the position, time features, and motion features of the trajectory are extracted from each trajectory data, and then the weighted kernel functions are fused through linear combination method. These kernel functions are applied to construct trajectory feature maps and using sparse subgraphs to detect abnormal trajectories; Wang et al. [6] proposed a method to consider the geospatial constraints of the trajectory and avoid the problem of sparsity, embed the geographic information and topological constraints of the trajectory into the structured vector space, and use RNN and CNN to model the general characteristics of the trajectory. This method can identify abnormal trajectories and determine which parts are the causes of abnormal trajectories; Zhang et al. [7] proposed an abnormal trajectory detection method combining emd DBSCAN algorithm and LSTM model with weighted loss function for ship trajectory data. The emd DBSCAN algorithm is used to cluster ship trajectories, identify ship trajectories with significant behaviors as anomalies, use LSTM model to train abnormal ship trajectory detection model, and increase the weight of abnormal trajectories in the loss function. The detection effect of this method is better than other algorithms; Wu et al. [8,9] propose a lightweight and authenticated key agreement protocol based on fog nodes in SIoV; they also propose a new provably secure authentication protocol to negotiate a session key before transmitting traffic information. Deng et al. [10] proposed a spatiotemporal graph convolutional adversarial network (STGAN) for anomaly detection, which designed a spatiotemporal generator to predict normal traffic conditions, a spatiotemporal discriminator to determine whether the input data is true, and combined the capabilities of the two detectors to detect abnormal traffic conditions. Real datasets were used for detection, achieving good detection results; Lai et al. [11] proposed a time-series outlier classification method based on behavior, and classified outlier into outlier and pattern outlier, and tested the method and obtained good detection results; Li et al. [12] proposed a comprehensive framework to study traffic prediction, data compression, and anomaly detection issues, and defined short-term trends to improve the prediction accuracy of short-term trend points and perform anomaly detection. Zhang et al. [13]

proposed a short-term traffic flow prediction algorithm of Quantum Genetic Alogrithm-Learning Vector Quantization(QGA-LVQ) neural network to forecast the changes of traffic flow.

In terms of data repair, Sun et al. [14] analyzed the repair effects of different interpolation methods, studied the repair effects of different linear regression multiple imputations, and provided different interpolation methods for different types of data; Zhang et al. [15] used a low rank proof interpolation method under alternating least squares to repair air quality monitoring data, and achieved better repair results compared to other methods; Zhang et al. [16] used the cubic spline method to interpolate and repair missing points between abnormal ship trajectories, improving the integrity and continuity of the trajectories; Zhao et al. [17] used an improved inverse distance weight interpolation method to repair the loss of information during vehicle driving. While obtaining a complete trajectory, it also has higher interpolation accuracy and is suitable for repairing vehicle trajectory data; Wang et al. [18] used spatiotemporal interpolation to repair missing data based on 3D shape functions, and compared it with traditional interpolation methods. The results showed that this method has higher accuracy in repairing missing traffic flow data; Meng et al. [19] studied the repair methods of different interpolation methods on traffic flow data, analyzed the results of different data, and found that using spatiotemporal correlation interpolation method has good results in repairing traffic flow data; Qiao et al. [20] proposed a K-nearest neighbor interpolation subspace clustering algorithm for high-dimensional feature missing data to address the problem caused by missing data features. The results showed that this algorithm can effectively handle high-dimensional feature missing data; Gao [21] obtained the optimal result by performing multiple imputations on missing data and conducting error analysis on the imputed results, achieving the repair effect on missing data; Deng et al. [22] analyzed the causes of different missing data and the application fields of different interpolation methods, providing the applicable fields for different types of data; Chen [23] studied the absence of time series and provided conclusions on the application of different interpolation methods in time series, analyzing the advantages and disadvantages of different interpolation methods in time series. The structure of this article is as follows. The second part introduces the anomaly detection process based on ETC data. The third part introduces different methods for repairing abnormal trajectories. The fourth part conducts experiments and analyzes the experimental results. The fifth part summarizes this article.

2. **Related Work.**

ETC transaction data belongs to a type of transportation big data, which includes transaction information, vehicle information, etc. Deep mining of ETC transaction data can obtain the distribution of traffic flow on road sections, as well as the driving speed of vehicles on each road section. This section uses ETC data provided by a high-speed company in a certain southeastern province. There are a total of 1091 gantry frames in the province, including 1021 normal gantry frames and 70 provincial boundary gantry frames, 64 virtual gantry frames. This section analyzes the basic information contained in ETC trading data, using a day's trading data from the province for research, and identifies issues in ETC trading data.

2.1. **Related definitions.**

ETC transaction data EData: The vehicle passing through the ETC gantry will generate a transaction record EData for the vehicle, which includes 103 business fields. The key fields are mainly vehicle identification (Card Plate, VehPlate, OBUPlate, OBUID), gantry data (FlagID, FlagName), transaction time (TradeTime), etc., as shown in TABLE 1.

Table 1. ETC Transaction Data (EData) Part Attribute Table (Desensitization)

Num	Name	Examples
1	TRADETIME	2021/06/02 20:00:01
2	FLAGID	3502**
3	FLAGTYPE	1
4	FLAGNAME	**to**
5	OBUID	66AD40**
6	PASSID	01*******423
7	OBUPLATE	********
8	ENTIME	2021/06/02 7:48:39
9	ENSTATION	350446**
10	VEHCLASS	1

Data anomaly rate: In the process of recording data, there are recording anomalies that do not meet the standards, which we call abnormal data. The calculation formula for abnormal data of different dimensions is expressed as:

$$ERate_{name} = \frac{ErrorNum_{name}}{TotalNum_{name}} \times 100\% \tag{1}$$

Among them, $ERate_{name}$ represents the anomaly rate of the data in this dimension, mainly including OBUPlate, VehPlate, CardPlate, etc. $ErrorNum_{name}$represents the amount of incorrect data in this dimension, $TotalNum_{name}$ represents the total amount of data for this dimension.

According to statistics, the province currently generates an average of 5-7 million transaction data per day, with holidays exceeding 10 million. The rich dimensions of ETC trading data provide a multidimensional perspective for data analysis, but its shortcomings, garbled code, or errors cannot be ignored.

2.2. Abnormal trajectory detection process.

Based on the preliminary preparation work, the corresponding OD pairs extracted and the corresponding OD path library are used to detect the vehicle trajectory extracted from the fusion of ETC transaction data and toll station data. The process of anomaly detection is shown in FIGURE 1.

Firstly, the ETC data and toll station data are fused using a fusion algorithm to obtain transaction data including the entrance and exit of the toll station. Then, vehicle trajectories are constructed based on the ETC transaction data, and the transaction trajectories of each vehicle are extracted. Compare the length of the trajectory with the corresponding length in the OD path library. If it is consistent, compare whether each gantry is equal. If it is equal, the trajectory is a normal trajectory. If not equal, then the trajectory is an abnormal trajectory; Return abnormal points in the detected trajectory and analyze the abnormal situation.

3. Research on anomaly repair.

Based on the analysis of abnormal trajectories in the above chapters, we can see that there are the least occurrences of duplicate transactions, followed by missed transactions, and the most frequent occurrences are missed transactions. In the repeated transaction trajectory, we only need to delete the repeated transaction data, and then verify whether the trajectory composed of the remaining data is composite with the highway traffic model. If it matches, it is the correct trajectory and transaction data. If it does not match, abnormal trajectories will be detected, and subsequent repair work will be carried out. For incorrect trading trajectories, they are

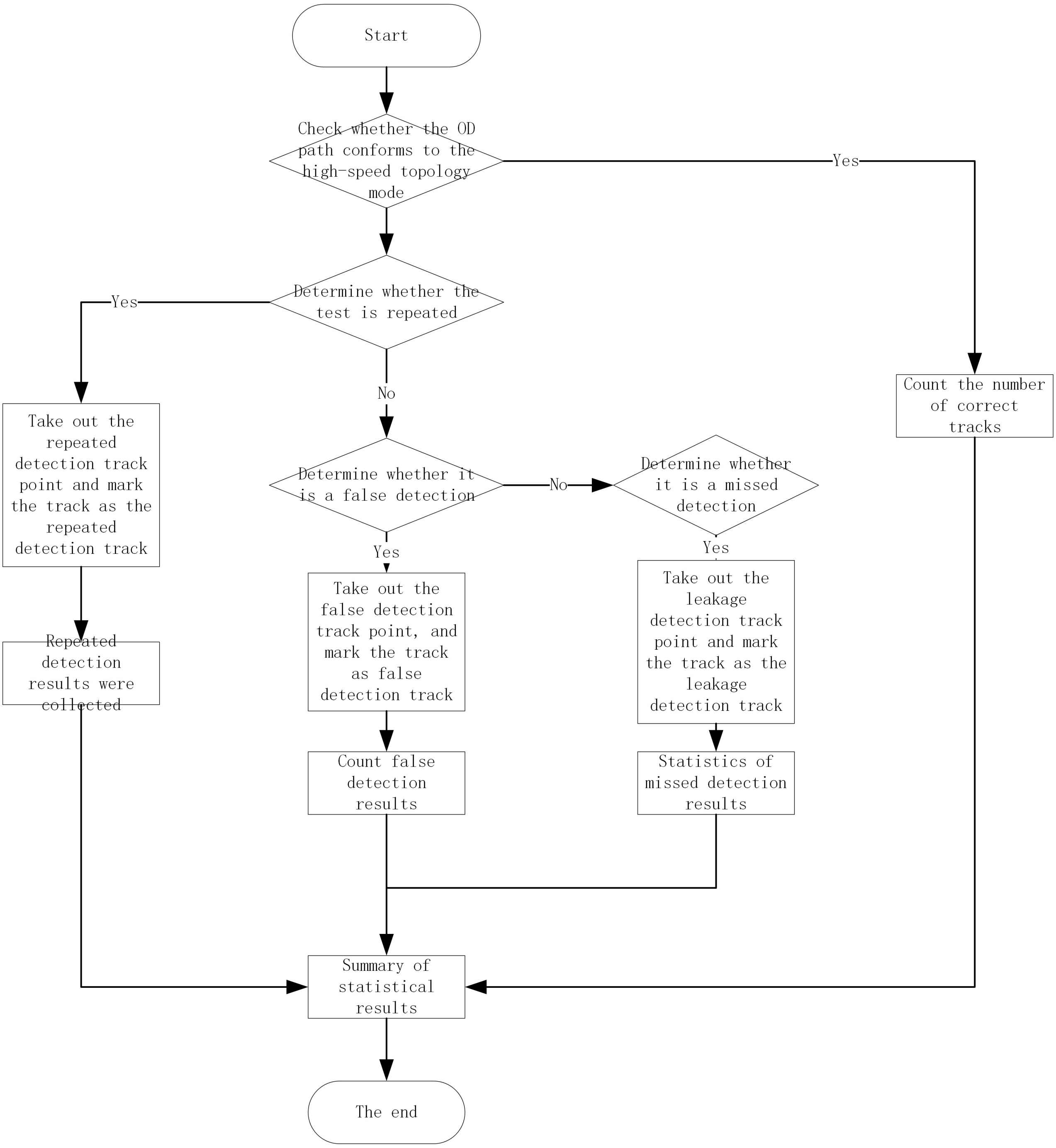
Start
Check whether the OD path conforms to the high-speed topology mode
Yes
Determine whether the test is repeated
Yes
No
Take out the repeated detection track point and mark the track as the repeated detection track
Determine whether it is a false detection
No
Determine whether it is a missed detection
Count the number of correct tracks
Yes
Yes
Repeated detection results were collected
Take out the false detection track point, and mark the track as false detection track
Take out the leakage detection track point and mark the track as the leakage detection track
Count false detection results
Statistics of missed detection results
Summary of statistical results
The end

Figure 1. Speed and Traffic Flow Time-Variation

inherently incorrect trading events, so we directly delete them and then detect the trading trajectories formed by the remaining data. However, the above method is not applicable for missing transactions. For missing transaction data, due to the lack of completeness of its trajectory, we need to reconstruct the transaction data of its missing points. Therefore, the repair of missing transaction trajectories mainly focuses on repairing the missing gantry transaction data in the missing transaction trajectories. After the analysis in the previous chapters, we know that ETC transaction data contains numerous dimensional information, including transaction time, travel ID, gantry name, gantry ID, and relevant identification information of passing vehicles. Among these data, the travel ID is the same for a trajectory. The name and ID of the gantry can be used to detect the missing gantry and obtain the corresponding gantry information based on the gantry list. The relevant identification information of passing vehicles can be obtained from other normal transaction information in the trajectory; But the missing transaction time data needs to be repaired according to relevant methods. The following is an analysis and research on methods for repairing missing data.

3.1. **A Mean Based Data Restoration Method.** In the field of data science, data quality is crucial, but missing values and outlier in data will affect our analysis and modeling of data. Therefore, data repair is necessary, and statistical data repair methods are one of the most common data repair methods. One of the most common statistical based repair methods is mean padding, which uses the mean or median of the entire dataset to replace missing values. We can fill in the missing transaction time in ETC transaction data by using the average travel time of the missing road section.

$$\overline{y} = \frac{\sum\limits_{i=1}^{n_1} y_i}{n_1} \tag{2}$$

Among them, $\bar{y}$ is the mean of the sample dataset, y_i represents the data value, n_1 represents the capacity of the missing data in the sample dataset, and n represents the capacity of the sample dataset.

When using mean interpolation, the estimated overall mean after interpolation is shown in Formula (2)

$$\hat{\overline{Y}} = \frac{1}{n}\left[\sum_{i=1}^{n_1} y_i + n_1\bar{y}\right] \tag{3}$$

Mean interpolation is one of the most common interpolation methods. Although it can meet the requirement of unbiased estimation in terms of overall characteristics, using only a single numerical interpolation can lead to distorted dataset distribution, reduced overall variance, and missing fluctuation information in the interpolated data. However, mean interpolation is the most common method in practical research applications. In cases of low loss rates, mean interpolation is considered a convenient and fast interpolation method. Therefore, it can be used as a comparison method to repair missing transaction times in ETC transaction data. In the repair of ETC transaction data, the mean of the travel time of vehicles of the same type on the same road segment in historical data is used for mean interpolation.

The specific algorithm for repairing the passage time in the missing trajectory of ETC is as follows in TABLE 2 :

This algorithm first groups historical trajectory data according to different road sections, and then calculates its travel time according to different vehicle types. Secondly, extract the corresponding road section where the missing transaction gantry is located, as well as the corresponding vehicle type, and match the travel time of different vehicle types on different road sections of historical data. Then, use this time as the travel time of the missing transaction section in the missing transaction trajectory. Finally, combine the trading time of the previous gantry before the missed trading gantry to repair the trading time of the missed trading gantry.

3.2. **A Repair Method Based on the Average Speed of Vehicles.** When a vehicle passes through the ETC gantry, transaction records are generated. At the same time, we can determine the distance it passes through based on topological relationships, and

TABLE 2. Mean interpolation algorithm

```
Algorithm: Mean interpolation algorithm
Input:  data1(Complete trajectory data),data2(Missing trading gantry trajectory data)
Outputre_data2(Fix transaction data with missing transaction time)
Step1:  passtime=data1. gropuby (flagid)/* Grouping the travel time of complete data*/
Step2:  vehclass_passtime = mean(gropupby(vehclass))/*Calculate the average travel time for different vehicle types*/
Step3:  for data in data2:/*Traverse missing transaction trajectories*/
Step4:  if (data2.flagid in data1) and (data2.vehclass in data1)
Step5:  re_time = vehclass_passtime /*Retrieve the average travel time of the corresponding vehicle type on the corresponding road section*/
Step6:  re_data = pre_tradetime + re_time/*Use the trading time of the previous gantry in the trajectory to repair the current missing gantry trading time*/
Step7:  else:
Step8:  return re_data/*If there is no corresponding data for the road section data, the data will be directly returned*/
```

then calculate the average speed of the vehicle passing through the road section. Then, the average speed is used to calculate the missing transaction time. The calculated time is used as the missing passage time for interpolation, but this method needs to know the transaction time of the front and rear frames of the missing frames, as well as the distance between the front and rear frames. At the same time, the more missing frames there are, the greater the cumulative error calculated by this method, which will affect the accuracy of passage time restoration. The formula for calculating the average speed is as follows:

$$V = \frac{Dis_{before} + Dis_{after}}{time_{after} - time_{before}} \tag{4}$$

Among them, $time_{after}$ represents the transaction time of one gantry after the missing gantry, $time_{before}$ represents the transaction time of the previous gantry before the missing gantry, Dis_{total} represents the distance of the road section, Dis_{before} represents the distance from the missing gantry to the previous gantry, Dis_{after} represents the distance from the missing gantry to the next gantry, and V represents the distance passing through the section. After calculating the average speed of the road section with missing trading gantry, combined with Dis_{before} calculates the travel time from the missing gantry to the previous gantry to calculate the missing transaction time.

This algorithm first calculates the travel time of the vehicle passing through the section before and after the missing gantry based on the transaction time of the vehicle passing through the missing gantry. Then, based on the average speed calculation formula, the average speed of the vehicle is calculated. Then, calculate the passage time based on the distance from the previous gantry to the missing gantry; finally, calculate the transaction time of the vehicle passing through the missing gantry as the repair transaction time. The specific algorithm is shown in TABLE 3 below.

TABLE 3. Repair algorithm based on average speed

```
Algorithm: Repair algorithm based on average speed
Input:  data(Missing transaction trajectory data)
Outputre_data (Fixed transaction data)
Step1:  t1 = data.flag1[before_tradetime]/*The trading time of the previous
        gantry before the missed trading gantry*/
Step2:  t2 = data.flag3[after_tradetime]/*The trading time of the last
        gantry after the missed trading gantry*/
Step3:  passtime = t2 – t1/*Calculate the vehicle travel time in the missed
        transaction section*/
Step4:  dis1 = get_dis(flag1,flag3)/*Obtain the distance of the section*/
Step5:  dis2 = get_dis(flag1,flag2) /*Obtain the distance from the missing
        gantry to the previous gantry*/
Step6:  v = dis1/(t/3600) /*Calculate the average speed of vehicles on the
        road section*/
Step7:  data.re_tradetime = t1 + (dis2 /v)/*Calculate the transaction time
        of the vehicle passing through the missing gantry*/
Step8:  return data /*Returns the repaired trajectory data*/
```

3.3. **Restoration algorithm based on random forest.** Machine learning based data repair methods have attracted much attention in recent years. Compared to traditional statistical methods, machine learning based methods usually have stronger adaptability and generalization performance. Common machine learning methods include regression based method, clustering based method, generative model based method, and semi supervised learning based method. Among them, the random forest algorithm is one of the most powerful and commonly used supervisory algorithms. As an extended variant of the parallel algorithm Bagging, it has the ability to solve both classification problems and regression problems. When you are not sure what method to use for the problem to be solved, random forest will be a better choice. On the basis of decision tree, random forest constructs a more powerful classification or regression model by combining multiple decision trees. Compared with a single decision tree, the advantage of random forest is that it can avoid overfitting and improve generalization ability. Moreover, random forest also performs well in processing high-dimensional data. Random forest algorithm is suitable for dealing with all kinds of problems. In essence, it is an integrated learning method. Integrated algorithms can predict the final result by integrating multiple algorithms and synthesizing the results of multiple models, which can achieve better results than a single model. The steps of random forest algorithm are shown in the following FIGURE 2:

Firstly, by randomly sampling dataset D, multiple training subsets are obtained. Then use subset data for training to establish decision trees, and randomly select K features from the selection to make each decision tree grow as long as possible without pruning during the process. Finally, after obtaining the required number of decision trees, calculate the average of their outputs as the prediction result.

Hyperparameter setting is an important work before using random forest for learning and training. If the parameter setting is unreasonable, it will lead to overfitting and under fitting. There are currently two commonly used setting methods: the first is the grid parameter optimization method. By setting the grid parameters, the algorithm will

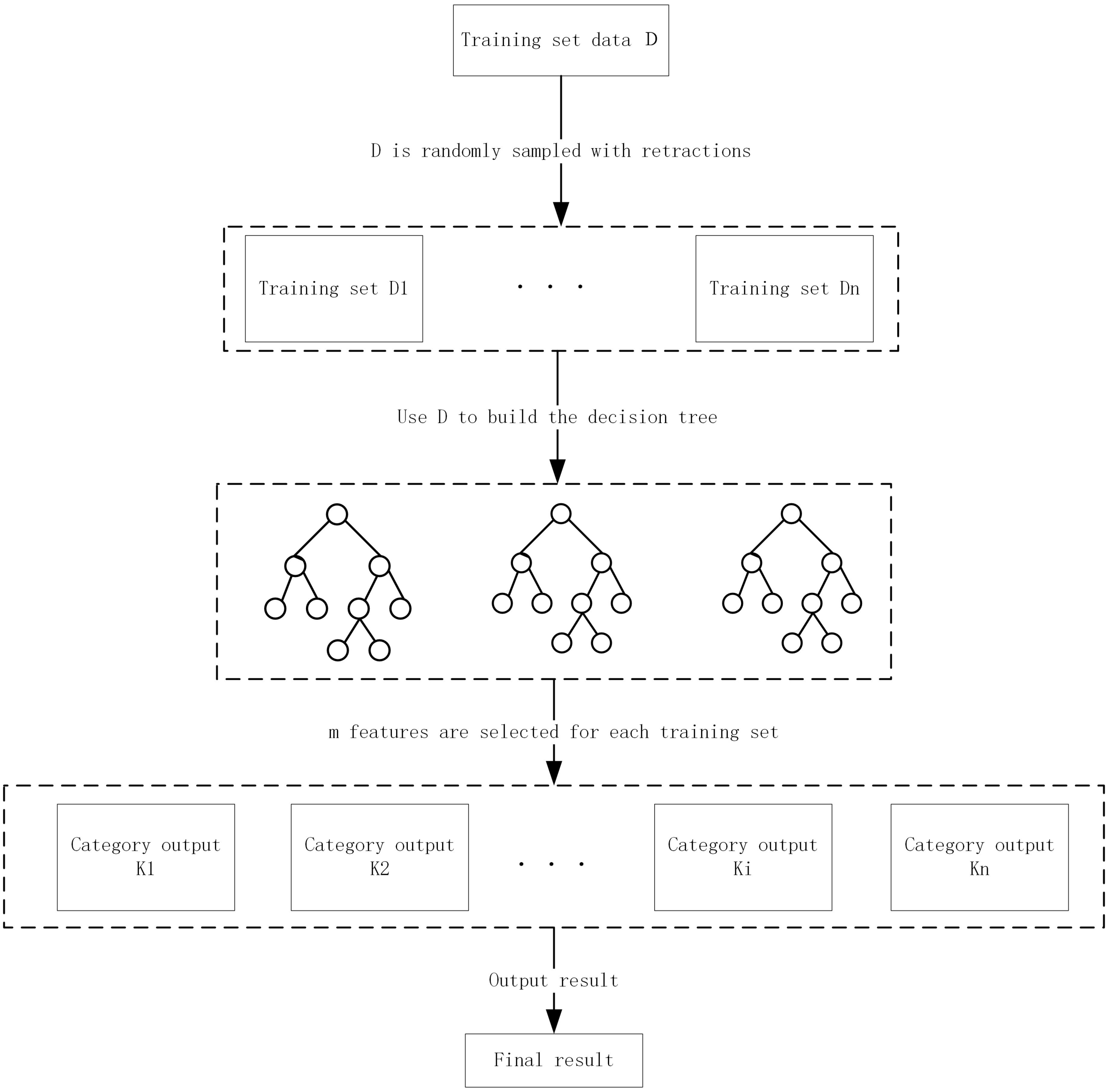

FIGURE 2. Speed and Traffic Flow Time-Variation

automatically select parameters within the grid according to the specified compensation to obtain the optimal parameters. Another approach is the experiential approach, which sets parameters based on personal experience and fine-tuning based on training results. In this paper, the grid parameter optimization method is selected to set hyperparameter.

Support Vector Machine Regression (SVR) is a nonlinear regression model suitable for various types of data, including continuous and discrete data. It uses kernel functions to map data to high-dimensional space, and then constructs a linear regression model in the high-dimensional space. By training this model, the output values of new data points can be predicted. This article selects this model as a comparative model for comparative experiments.

The eigenvector and data preprocessing results are obtained through the above chapters, and the passage time of the section is predicted using the random forest algorithm, and then the missing transaction time is repaired using the predicted passage time. First of all, data preprocessing is carried out. Secondly, the original sample set consisting of vehicle transaction section data set is split into training set and test set according to the ratio of 7: 3. Then, the model construction and parameter optimization are carried out to build a random forest model with section characteristics to predict the travel time, and repair the transaction time of the missing gantry through the predicted travel time.

4. **Experimental results.**

4.1. **Abnormal detection results.**

According to the proposed detection process, the trajectory of the vehicle to be detected is detected. The main detection results are divided into duplicate detection, false detection, missed detection, etc.

Duplicate detection trajectory refers to the occurrence of duplicate gantry information in the transaction trajectory, including duplicate gantry IDs, duplicate transaction times, duplicate passids, duplicate obuids, etc., where two identical pieces of data appear simultaneously in the same itinerary. Duplicate detection may be due to the RSU detection equipment of the gantry detecting information about the same vehicle at a similar time and transmitting it to the backend. During the detection period, multiple RSU devices on a single gantry share a gantry ID, resulting in multiple RSU devices simultaneously recording transactions, resulting in duplicate transaction data in ETC transaction data. The following definition is made to calculate the repetition rate of the trajectory:

$$duptraj_{rate} = \frac{duptraj_{num}}{traj_{totalnum}} \tag{5}$$

Among them, $duptraj_{rate}$ represents the repetition rate of the trajectory, $duptraj_{num}$ is the trajectory containing duplicate transaction data, $traj_{totalnum}$ is the number of all trajectories. A total of 714 trajectories were detected with duplicate transactions in all journeys. From FIGURE 3, we can see there are 644 repetitive trajectories in the first type of trajectory (Class I trajectory), 29 repetitive trajectories in the second type of trajectory (Class II trajectory), 24 repetitive trajectories in the third type of trajectory (Class III trajectory), and 17 repetitive trajectories in the fourth type of trajectory (Class IV trajectory). The specific situation is shown in the following figure. The repetition rate of all trajectories is 0.009%, which is very low.

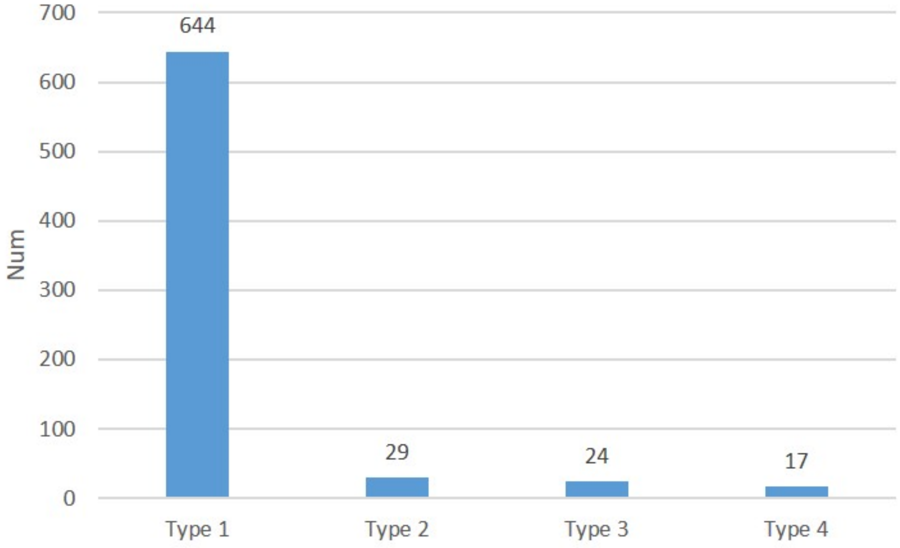

FIGURE 3. Number of repeated trajectories in different trajectories

Misdetection trajectory refers to the nodes that appear in the transaction trajectory and should not appear in the topology relationship, that is, the wrong gantry nodes, most of which are opposite gantry nodes, are mistakenly detected in the current transaction trajectory. The definition of false detection rate is as follows:

$$error_{rate} = \frac{errortraj_{num}}{traj_{totalnum}} \tag{6}$$

Among them, $error_{rate}$ represents the error detection rate of the trajectory, $error_{rate}$ represents the number of trajectory errors detected, $traj\_total\_num$ is the number of all trajectories. The false detections contained in different trajectories in the four types of trajectories are as follows: 4606 false detections in the first type of trajectory, 30 false detections in the second type of trajectory, 37 false detections in the third type of trajectory, 21 false detections in the fourth type of trajectory, and a total of 4694 false detections in all trajectories. The overall false detection rate is 0.59%, and the specific false detection situation is shown in the following FIGURE 4.

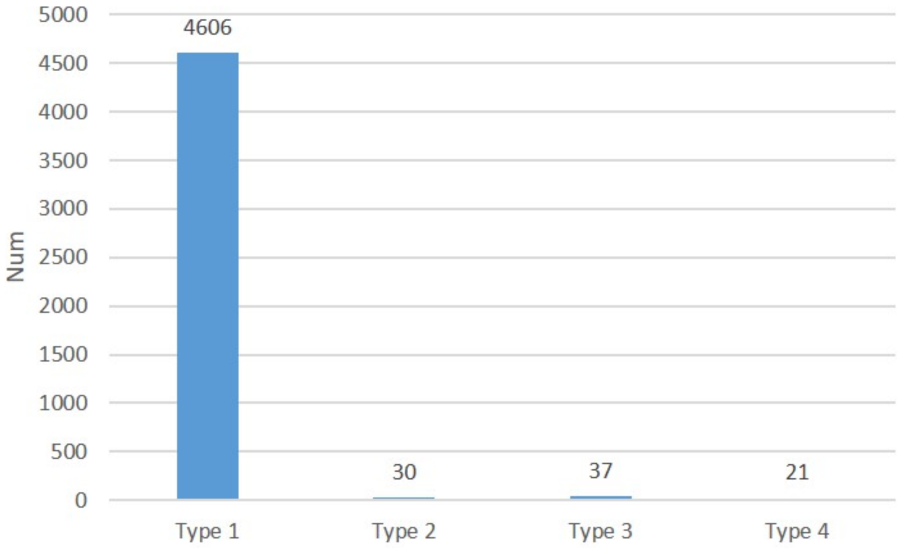

FIGURE 4. Number of incorrect trading trajectories in different trajectories

Missing detection trajectory refers to the occurrence of missing gantry frames in the transaction trajectory, resulting in incomplete transaction topology of the vehicle. However, filling in the missing gantry frames can form a complete trajectory. The definition of missed detection rate is as follows:

$$miss_{rate} = \frac{misstraj_{num}}{traj_{totalnum}} \tag{7}$$

Among them, $miss_{rate}$ represents the missed detection rate of the trajectory, $misstraj_{num}$ represents the number of trajectories that generated missed transactions, $trajtotal_{num}$represents the number of all trajectories. Among the four types of trajectories, there are 85822 missed detection trajectories in the first type, 6540 missed detection trajectories in the second type, 3439 missed detection trajectories in the third type, and 4064 missed detection trajectories in the fourth type, totaling 99865 missed detection trajectories. The missed detection trajectory rate is 12.5%. The specific false detection situation is shown in the following FIGURE 5.

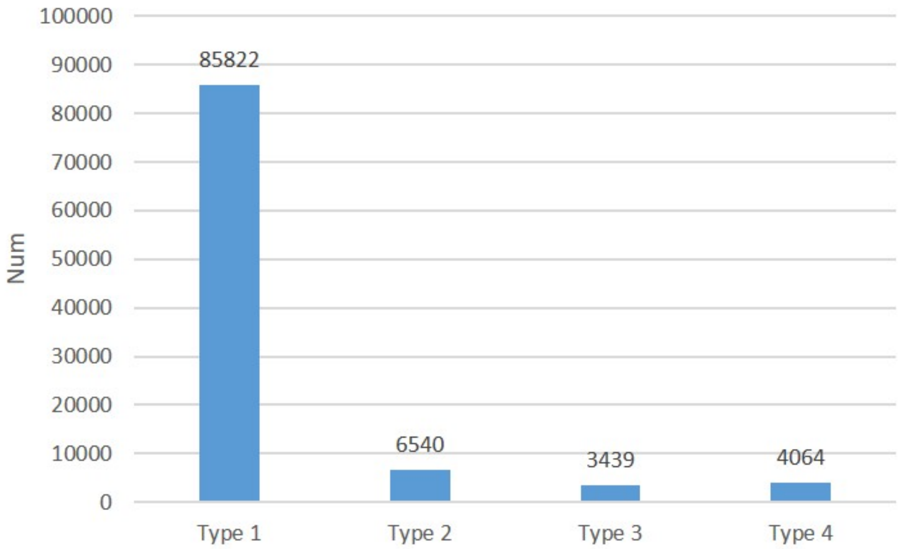

FIGURE 5. Number of missed trading trajectories in different trajectories

In addition, there is a portion of trajectories that include both false detection and missed detection gantry frames. There are 25311 trajectories in the first category, 175 trajectories in the second category, 302 trajectories in the third category, and 593 trajectories in the fourth category, totaling 26831 trajectories, accounting for 3.3%.

4.2. **Data repair results.** The evaluation indicators used include mean absolute error (MAE) and root mean square error (RMSE). The MAE indicator can reflect the actual situation between imputed values and missing values, and the smaller its value, the higher the accuracy of imputed values. The specific calculation formula is as follows:

$$MAE = \frac{1}{m}\sum_{i=1}^{m} \mid \widehat{y_i} - \ y_i \mid \tag{8}$$

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m} (y_i - \widehat{y_i})^2} \tag{9}$$

The specific results are shown in the following FIGURE 6. Method 1 is based on statistical methods to repair the missing transaction time, Method 2 uses the average speed of vehicles passing through the road section to calculate the missing transaction time, Method 3 uses SVR in machine learning methods for repair, and Method 4 uses random forests in machine learning for repair.

The road sections from Section 1 to Section 4 do not include service areas. According to the results, random forest has the best restoration effect among the four methods. For MAE, the effect of SVR in Section 3 is slightly better than that of random forest, but in other sections, the performance of random forest is the best. For RMSE method 2, the effect is better, followed by method 4, and for methods 1 and 2, the performance in RMSE is poor. Comprehensively, the random forest method performs best on the road section without service area, and can be used as a data repair method to repair the missing transaction track.

Sections 5 to 8 include the service area section, and the location of the service area is within the section before the missing gantry. From the FIGURE 7, we can see that except for random forest, the error of other methods of MAE is large, and the error of the

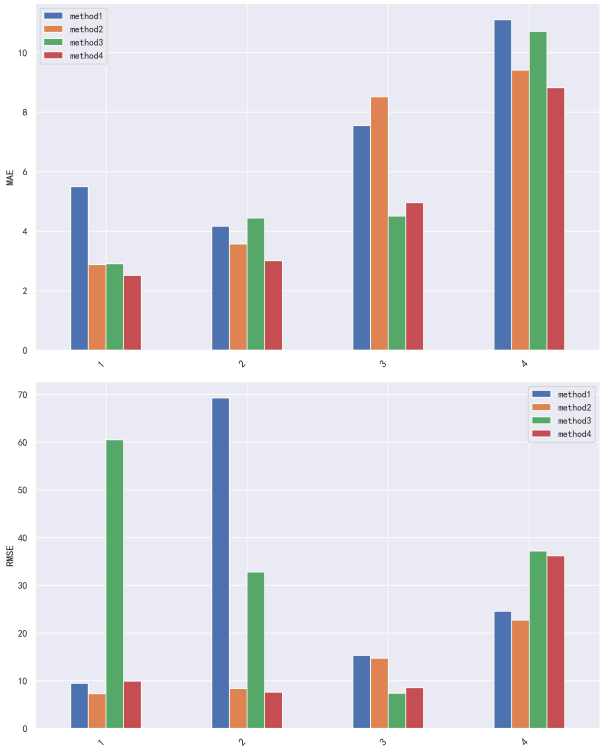

Figure 6. Comparison results of different repair methods for sections 1-4

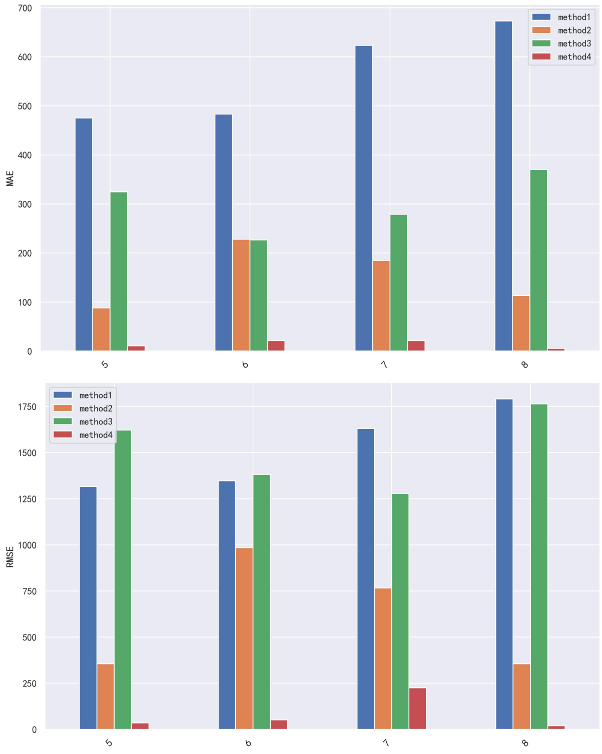

Figure 7. Comparison results of different repair methods for sections 5-8

average interpolation method is the largest, followed by the error of the average speed of the road section. The result of RMSE random forest is also far better than other methods, so random forest can also be used to repair missing data of road sections including service areas.

According to the repair results of the experimental data, the MAE of the proposed random forest repair algorithm based on segment characteristics is 4.82s in the overall repair results of the road sections without service areas, while the overall repair effect of the road sections with service areas is poor, with the MAE of 14.77s. According to the FIGURE 6 and FIGURE 7, the performance of the random forest algorithm based on segment characteristics is better than other methods, Therefore, this method can be used to repair missing transaction trajectories in ETC transaction data.

5. **Conclusion.** This article analyzes the problems with the trajectory of highway vehicles and proposes a trajectory anomaly detection method for detecting outliers in trajectory data. Then, combined with the feature database construction algorithm to build the characteristics of the provincial road sections, a method based on random forest algorithm to repair the missing transaction track is proposed, and the repair results are added to the ETC transaction data to improve the integrity of the transaction data. However, the detection algorithm proposed in this article has poor detection performance when there are indeed too many gantry frames in the transaction trajectory, and there is still room for improvement. In terms of trajectory repair, some car owners on highways engage in fee evasion behavior, which can also lead to abnormal trajectories. Moreover, such abnormal trajectories often have a large amount of missing transaction data, making trajectory repair difficult. This is also an area for future research expansion.

Acknowledgment. This work is partially supported by the Renewable Energy Technology Research institution of Fujan University of Technology Ningde, China (Funding number:KY310338), the 2020 Fujian Province “Belt and Road” Technology Innovation Platform (Funding number:2020D002), the Provincial Candidates for the Hundred, Thousand and Ten Thousand Talent of Fujian (Funding number:GY-Z19113), the Patent Grant project (Funding number:GY-Z18081, GY-Z19099, GY-Z20074), Horizontal projects (Funding number:GY-H-20077), Municipal level science and technology projects (Funding number:GY-Z-22006, GY-Z-220230).

The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] Y. Wang, F. Gao, and L. Cao, “TRASMIL: A local anomaly detection framework based on trajectory segmentation and multi-instance learning,” *Computer Vision and Image Understanding*, vol. 117, no. 10, pp. 1273-1286, 2013.

[2] M. Braei, and S. Wanger, “Anomaly detection in univariate time-series: A survey on the state-of-the-art,” arXiv preprint arXiv:2004.00433, 2020.

[3] G. Pang, C. Shen, L. Cao, and A. Van, “Deep learning for anomaly detection: A review,” *ACM Computing Surveys (CSUR)*, vol. 54, no. 2, pp. 1-38, 2021.

[4] J. Hu, X. Zhang, and S. Maybank, “Abnormal driving detection with normalized driving behavior data: a deep learning approach,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 7, pp. 6943 - 6951, 2020.

[5] X. Zhao, Y. Rao, J. Cai, and W. Ma, “Abnormal trajectory detection based on a sparse subgraph,” *IEEE Access*, vol. 8, pp. 29987-30000, 2020.

[6] H. Wang, J. Feng, L. Sun, K. An, and H. Chai, “Abnormal trajectory detection based on geospatial consistent modeling,” *IEEE Access*, vol. 9, pp. 184633-184643, 2020.

[7] T. Zhang, S. Zhao, B. Cheng, and J.-L. Chen, “ATeDLW: Intelligent Detection of Abnormal Trajectory in Ship Data Service System,” *2021 IEEE International Conference on Services Computing (SCC)*, IEEE, 2021, pp. 401-406.

[8] T.-Y. Wu, X.-L. Guo, L.Yang, Q. Meng, and C.-M. Chen, “A lightweight authenticated key agreement protocol using fog nodes in Social Internet of vehicles,” *Mobile Information Systems*, vol. 2021, 3277113, 2021.

[9] T.-Y. Wu, Z.-Y. Lee, L. Yang, and C.-M. Chen, “A Provably Secure Authentication and Key Exchange Protocol in Vehicular Ad Hoc Networks ,” *Security and Communication Networks*, vol. 2021, 9944460, 2021.

[10] L. Deng, D. Lian, Z. Huang, and E.-H. Chen, “Graph convolutional adversarial networks for spatiotemporal anomaly detection,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 6, pp. 2416-2428, 2022.

[11] K.-H. Lai, D. Zha, J. Xu, Y. Zhao, G.-C Wang, and H. Xia, “Revisiting time series outlier detection: Definitions and benchmarks,” *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (round 1)*, 2021, pp. 58-65.

[12] L. Li, X. Su, Y. Zhang, J. Hu, and Z. Li, “Traffic prediction, data compression, abnormal data detection and missing data imputation: An integrated study based on the decomposition of traffic time series,” *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, IEEE, 2014, pp. 282-289.

[13] F.-Q. Zhang, T.-Y. Wu, Y.-O. Wang, R. Xiong, G.-Y. Ding, P. Mei, and L.-Y. Liu, “Application of quantum genetic optimization of LVQ neural network in smart city traffic network prediction,” *IEEE Access*, vol. 8, pp. 104555-104564, 2020.

[14] L.-L. Sun, S.-J. Dong, and G.-J. Yang, “Selection of Interpolation Multiplicity for Common Multiple Interpolation Methods,” *Statistics & Decision*, vol. 35, no. 23, pp. 5-10, 2019.

[15] B. Zhang, and G.-J. Song, “An Empirical Study on the Methods for Handling Missing Data in Large Scale Air Quality Monitoring,” *China Environmental Science*, vol. 42, no. 25, pp. 2078-2087, 2022.

[16] L.-X. Zhang, Y. Zhu, and W. Lu, “Research on Ship Trajectory Restoration Method Based on AIS Data,” *Journal of Northwestern Polytechnical University*, vol. 39, no. 01, pp. 119-125, 2021.

[17] Z.-X. Zhao, R.-T. Qu, and J.-W. Wang, “Vehicle trajectory reconstruction method based on improved inverse distance weight interpolation,” *Journal of Highway and Transportation Research and Development*, vol. 35, no. 10, pp. 133-139, 2018.

[18] W. Wang, Z.-Y. Cheng, M.-Y. Liu, and Z.-S. Yang, “A method for repairing traffic flow fault data based on spatiotemporal correlation,” *Journal of Zhejiang University*, vol. 51, no. 09, pp. 1727-1734, 2017.

[19] H.-C. Meng, and S.-Y. Chen, “Comparative analysis of data processing methods for missing traffic flow data,” *Journal of Transport Information and Safety*, vol. 36, no. 02, pp. 61-67, 2018.

[20] Y.-J. Qiao, X.-L. Liu, and L. Bai, “K-nearest neighbor interpolation subspace clustering algorithm for high-dimensional feature missing data,” *Journal of Computer Application*, vol. 42, no. 11, pp. 3322-3329, 2022.

[21] J.-X. Gao, “Research on Data Restoration and Prediction Methods for Intelligent Transportation,” Master. dissertation, Tianjin University of Technology, 2020.

[22] J.-X. Deng, L.-B. Shan, D.-Q. He, and R. Tang, “Processing Methods and Development Trends of Missing Data,” *Statistics & Decision*, vol. 35, no. 23, pp. 28-34, 2019.

[23] Y.-S. Chen, “Overview of processing methods for missing data in time series,” *China Computer & Communication*, vol. 32, no. 10, pp. 19-22, 2020.